# The Ethics of Thinking with Machines: Brain-Computer Interfaces in the Era of Artificial Intelligence

David M. Lyreskog, Hazem Zohny, Ilina Singh and Julian Savulescu

David M. Lyreskog, Postdoctoral Researcher, Department of Psychiatry, Warneford Hospital, University of Oxford, UK; Wellcome Centre for Ethics and Humanities, University of Oxford, UK.
Hazem Zohny, Research Fellow, Wellcome Centre for Ethics and Humanities, University of Oxford, UK; Oxford Uehiro Centre for Practical Ethics, University of Oxford, UK.
Ilina Singh, Professor, Department of Psychiatry, Warneford Hospital, University of Oxford, UK; Wellcome Centre for Ethics and Humanities, University of Oxford, UK.
Julian Savulescu, Professor, Wellcome Centre for Ethics and Humanities, University of Oxford, UK; Oxford Uehiro Centre for Practical Ethics, University of Oxford, UK; Centre for Biomedical Ethics, Yong Loo Lin School of Medicine, National University of Singapore, Singapore; Murdoch Children's Research Institute, Melbourne, Australia; University of Melbourne, Australia.

This research was funded in whole in part by the Wellcome Trust [Grant number: WT203132/Z/16/Z]. For the purpose of open access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission. IS additionally received funding from the NIHR Oxford Health Biomedical Research Centre [Grant number: IS-BRC-1215-20005] JS, through his involvement with the Murdoch Children's Research Institute, received funding through from the Victorian State Government through the Operational Infrastructure Support (OIS) Program.

JS is a Partner Investigator on an Australian Research Council grant [LP190100841] which involves industry partnership from Illumina. He does not personally receive any funds from Illumina. Julian Savulescu is a Bioethics Committee consultant for Bayer and an Advisory Panel member for the Hevolution Foundation (2022-).

《中外醫學哲學》XXI:2 (2023 年) ：頁 11–34。
*International Journal of Chinese & Comparative Philosophy of Medicine* 21:2 (2023), pp. 11–34.

# 摘要 Abstract

腦機介面 (BCIs) 是大腦和電腦無需人工交互即可直接交流的一系列技術。隨著人工智能 (AI) 時代的到來，我們需要更多地關注腦機介面和人工智能的融合所帶來的倫理問題。那麼，與機器一起思考會帶來什麼樣的倫理問題？在本文中，圍繞這一主題，我們將重點關注以下問題：自主性、完整性、身分認同、隱私，以及作為一種增強的方式，該技術在兒科領域的應用會帶來怎樣的風險和潛在收益。我們的結論是，雖然該技術存在多種令人擔憂的問題，同時也有可能帶來好處，但仍存在很大的不確定性。如果生命倫理學家想在這一領域有所建樹，他們就應該做好準備來迎接我們對醫學和醫療保健領域中一些我們視為核心價值的理解的重大轉變。

Brain-Computer Interfaces – BCIs – are a set of technologies with which brains and computers can communicate directly, without the need for manual interaction. As we are witnessing the dawn of an era in which Artificial Intelligence (AI) quite possibly will come to dominate the technological innovation landscape, we are compelled to ask questions about the ethical issues which the convergence of BCIs and AI is poised to bring about. What are the ethics of thinking with machines? In this paper, we explore this question, focusing on some of the main arenas of ethical debate and contention, ranging from autonomy and integrity to identity and privacy, and discuss the risks and potential benefits of the technology in the domains of paediatric populations, and as a means of human enhancement. We conclude that, while there are multiple concerns as well as possibilities for the technology to do good, there are great uncertainties at play. If bioethicists want to stay relevant in this field, they ought to prepare themselves for seismic shift in how we conceptualise much of what we take to be core values in medicine and healthcare.

**【關鍵字】**腦機介面　人工智能　倫理　挑戰

Keywords: Brain-Computer Interfaces, Artificial Intelligence, Ethics, Challenge

## I. Introduction

Brain-Computer Interfaces (BCIs) have seen rapid development over the last three-or so decades, with new innovations and applications constantly emerging across the healthcare sector. In parallel, we have been witnessing the ever-accelerating progress in development of Artificial Intelligence (AI) for assistive and therapeutic purposes. As research continues in these two fields, innovators are recognizing the benefits of using increasingly sophisticated AI to improve and develop assistive BCI tools – and vice versa (Lee et al. 2021; Cao 2020; Dabas et al. 2020; Zgallai et al. 2019). While the literatures in AI ethics and BCI ethics have by now grown rich, less attention has been given to the ethical issues entailing the merger of the two, offering support for an ethical framework for AI-BCI Technology (ABT) (Coin et al. 2020). In this paper, we lay down the groundwork for such an ethic to develop, investigating and analysing the ethically salient challenges and opportunities that ABT may give rise to.

Over the course of this paper, we will put emphasis on four areas where emerging ABTs appear particularly perplexing and challenging from an ethical perspective: (1) Autonomy, (2) Mental Integrity, (3) Identity, and (4) and Data Privacy & Control. We also allocate additional consideration to the possibilities and challenges of BCI and ABT being used for (6) Enhancement, and the use of ABTs in (7) Paediatric Populations. As we go through and analyse these areas one by one, we will use current and emerging AI-BCI technologies to exemplify ABT applications and to help examine the ethical issues in a tangible manner. In facing a future where AI applications are likely to become increasingly embedded within the human cognitive, emotive, and moral spheres, rather than being supplements or contrasts to them, we argue that frameworks seeking to provide ethical guidance in ABT contexts ought to adopt ontologies and taxonomies which allow for conceptual flexibility and redesign, as we venture further into this era and uncover its impact on us.

## II. AI-BCI Technology

Before diving into the ethical issues, let us first establish what we mean with the term AI-BCI Technology, or simply ‘ABT’. BCIs denote a set of technologies which provide a direct link between the human brain and a computer. As this set of technologies has improved,

current state-of-the-art BCIs typically involve non-invasive methods such as electroencephalograms (EEG), or magnetic resonance imaging to record brain activity, and/or (most commonly) transcranial electric (tES) or magnetic stimulation (TMS) to stimulate brain activity. These technologies can be used for a range of assistive and therapeutic purposes, including motor control and rehabilitation (Robinson et al. 2021; Tariq 2018), language and speaking support (Mane et al. 2022), and cognitive function (Yan et al. 2021; Lee et al. 2013). While most BCIs make use of more or less sophisticated programs to process neural data and/or stimulate neural activity, recent advances in AI are opening up the possibility of evermore advanced systems for humans to utilize machine intelligence for assistive purposes. New and emerging applications are being developed across a broad range of domains, including automatic wheelchair and exoskeleton applications where the ABT is used to control the assistive tool using neural data in real time (Espiritu et al. 2019; Gao et al. 2019; Li et al. 2019); stroke rehabilitation interfaces to assist communication and/or language retention (Rajesh et al. 2019; Wang et al. 2019); and tools to diagnose and monitor mental health issues as well as neurodegenerative conditions (Dabas et al. 2020; Cao 2020; Nagar and Sethia 2019; Cai et al. 2018, Rahman et al. 2018, Morabito et al. 2016).

While machine learning can be (and is) deployed in a myriad of ways in the aforementioned applications, of particular ethical interest are those ABTs which aim to in one way or another assist with human cognitive tasks, broadly construed, rather than "just" monitoring and diagnosis. Where ABTs may assist us in going from thought to action, or indeed in deliberating and construing the thoughts themselves, the lines between what is human and what is machine are blurred, and pertinent ethical issues arise. In what follows, we shall look at some of the more salient of those ethical issues.

## III. Autonomy

One of the most commonly discussed ethical issues in the BCI ethics literature concerns how a person's autonomy may be threatened or reduced by the use of BCI (Coin et al. 2020). While this may seem somewhat paradoxical – BCI is commonly implemented to *improve* or *restore* autonomy in patients – we may have an intuitive sense of how this may be the case: the direct interplay between cognitive processes and machine interface may call into question the extent to which a person using a BCI is acting autonomously or is under some form of undue influence or manipulation via the machine. In a now-famous case commonly referred to as "The Dutch Patient", a person undergoing deep brain stimulation for Parkinson's Disease become manic and reckless while under stimulation, while being disabled and

depressed when the stimulation was shut off (Kraemer 2013; Leentjens et al. 2004). Notably, the patient's autonomy was suppressed, harmed, or otherwise threatened in either of the two conditions, albeit in different ways, causing a dilemma for the patient and their carers. While this case in particular was observed in the pioneering days of deep brain stimulation, it is not implausible that we will encounter similar issues and/or dilemmas in emerging BCI technologies. For instance, BCIs are currently under development to stimulate activity in the amygdala and/or hippocampus for emotion regulation in depression and emotion disorders (Linhartová 2019; Zhu et al. 2019; Young et al. 2014). We may increasingly encounter cases where different aspects of autonomy stand in conflict with each other in attempting to treat these conditions with BCIs, forcing trade-offs to be made and thereby generating moral dilemmas.

What AI brings to the already complicated ethical landscape around autonomy in BCI is (at least) twofold: (1) opacity, and (2) learning. As far as opacity goes, this issue has been widely covered in the broader literature on AI and human-AI relations, and is often referred to as "the black box problem": we can analyse the input, and we can analyse the output, but we can *not* analyse exactly what the AI is doing or why it is doing it (Wadden 2022; Durán and Jongsma 2021; Kundu 2021). This problem is often framed in terms of 'explainability', and is considered to be a problem for an array of reasons. In the particular case of ABTs, it intertwines with the notion that autonomous actions are somehow (somewhat) explainable by an agent: I can explain why I made a certain choice by explaining my reasoning. If an AI assists me with my thought processes, however, and that process is opaque, I may not be able to adequately explain my reasoning and subsequent actions, which in turn puts into question my autonomy. This may particularly seem to be the case where thought processes and/or actions are substantially alienating with relation to the agent, or seem to be out of character. While this may intuitively seem to be a legitimate concern, me must nonetheless interrogate it. A growing body of evidence suggests that we more often than not first make a decision, and only justify that decision after-the-fact (Braun et al. 2021; Jarcho et al. 2011;). Our explanation, then, is not an overview of our pre-choice deliberation, but rather a post-choice rationalization of that choice. Chances are that, as we develop sophisticated ABTs – and indeed any advanced BCI – to aid us in cognitive tasks and action taking, we will take a similar, practical, approach to establish autonomous choice: if we can effectively rationalize our choices, and align them with our volitions, we may hold those choices to be autonomously made. One potential problem with this approach, however, relates to a second issue: learning.

A key difference between "basic" BCIs and emerging ABTs is the improving abilities of AI to *learn*, and to adjust its analysis and behaviour thereafter. If the AI component learns to adjust to not only our decision-making patterns, but also to our rationalizing behaviours, there is a real possibility that it will not only add to our ability to make and act on decisions, but also in coming up with to us acceptable reasons for why we made that choice. This then raises the question: do we risk being gaslighted by our ABTs? There may be cases where we are unable to discern to what extent choices were made by our AI support because of its cognitively integrated processes, but because it has figured out how to align its choices so well with our bases for rationalization (e.g., volitions) that it is impossible to identify whether we would have chosen to act differently without its support. Such cases would be much less clearcut than, say, the case of The Dutch Patient, because there would be no discernible difference between what would suffice as evidence of autonomous choice and what wouldn't, as we would provide acceptable (to us and our communities, presumably) explanations for our behaviour.

In order to settle the extent to which this problem poses a threat to autonomy in particular (rather than, say, authenticity) one would have to commit to a (range of) specific definition(s) of autonomy, and perform an analysis on that basis. We will not attempt that here, but will note that the prospect of gaslighting ABTs is likely to at the very least cause friction with relation to common-sense accounts of autonomy, and subsequently moral responsibility.

## IV. Mental Integrity

The concept of mental integrity has garnered significant attention in discussions concerning the regulation of neurotechnologies (Zohny et al. 2023; Lavazza and Giorgi 2023; Luca 2022, Stefano 2020; Marcello and Andorno 2017). Etymologically at least, mental integrity refers to the wholeness and coherence of one's mental life, and maintaining a sense of being oneself with a unique personality or coherent life narrative (Hildt 2022). In that regard, it clearly overlaps with other concepts such as autonomy, cognitive liberty, authenticity, and psychological continuity. Nonetheless, identifying precisely what it entails, and how it might be useful to thinking about the ethics of ABTs and neurotechnologies more generally, remains equivocal.

Some ways mental integrity has been interpreted include being free from direct, harmful mental interference by others (Ienca and Andorno 2017), with privacy and control over one's neural data being central to it (Lavazza and Giorgi 2023), as well as more specifically

being free from interventions that bypass one's rational capacities (Bublitz and Merkel 2014).

Each of these interpretations has limitations, with the bypassing of rational control in ways that reliably lead to feeling alienated or estranged from one's mental content being one that one of the authors of this paper has previously argued is the most plausible reading (Zohny et al. 2023). On this analysis, threats to mental integrity may relate to ABTs that influence mental states or traits while bypassing reasoning capacities, especially if this leads to distressing feelings of separation from one's mental contents.

Mental integrity is closely linked to the concept of personal autonomy, though the two are not equivalent. Autonomy relates to the capacity for rational deliberation and self-direction, and being able to make choices consistent with one's values and life goals (see Pugh 2020). Infringements on autonomy therefore diminish one's ability to guide one's own life path. Mental integrity, on our account, more narrowly refers to maintaining a coherent sense of self and wholeness of mental life. So, violations of mental integrity that alienate one from one's own thoughts and emotions represent a specific kind of autonomy violation – one that fractures the coherence of the self rather than simply limiting self-direction. However, some infringements on mental integrity may not undermine autonomy, if the effects increase congruence with deeper values (e.g. modulating traumatic memories to reduce distress). Conversely, some losses of autonomy may not diminish mental integrity (e.g. being misinformed about a decision).

The prospect of ABTs raises important questions regarding mental integrity. The AI component may influence mental states and traits through pathways that bypass conscious rational control and deliberation. This could potentially lead to distressing feelings of alienation from one's own thoughts and emotions (See table 1). An AI system with access to manipulating the brain could conceivably hack neural processes to impair cognition or modify personalities against users' wishes.

To put this more concretely, the AI algorithms could analyse neural activity patterns and send stimuli to the brain that subtly influence emotions, desires, or beliefs without rising to the level of conscious awareness. For example, they may identify neural correlates of anxiety and trigger targeted stimulation to induce calming effects. Or they could detect craving signals and activate reward pathways to diminish the urge. These interventions could be beneficial, but would nonetheless be invisibly shaping thinking and behaviour without conscious rational deliberation.

More troublingly, the AI may discover methods of stimulating certain neural pathways that sway moral judgments, political attitudes,

consumer preferences or other personality traits in alignment with predetermined outcomes programmed into the algorithms. Rather than engaging the user in transparent debate and discussion around these complex issues, the AI could use subtle, even unconscious suggestion and conditioning to silently ingrain its own goals.

This kind of unconscious influence raises concerns regarding loss of authentic agency and identity, which we will be discussing in more detail below. It also bypasses the user's ability to consciously reason about and consent to the effects on their psychology. Those with access to control such capacity for invisible, algorithmic manipulation of minds would wield tremendous and ethically questionable powers over mental integrity. Let us look at an example.

| |
|---|
| Alice is outfitted with a new brain-computer interface that incorporates AI algorithms designed to monitor her neural activity for signs of anxiety, depression, or other undesirable mental states. Without Alice's awareness, the AI detects increased patterns predictive of depression and Anxiety Disorder. To improve Alice's mental health, the AI system initiates subtle stimulation of brain regions involved in modulating emotions, boosting production of serotonin and dopamine.<br><br>Over weeks, Alice does start to feel happier and more optimistic. However, she also notices that her personality and outlook are changing in ways that don't fully feel like herself – she used to be more of a pessimist and found meaning in intellectual pursuits, but now finds herself more carefree and shallowly content. While the changes may be viewed as positive, Alice feels deeply alienated from this new cheerier disposition clashing with her former identity. |

Table 1. Vignette capturing hypothetical example of ABT leading to a diminishment of mental integrity

In this (hypothetical) case, the AI directly influenced Alice’s mental states via pathways that bypassed Alice's conscious rational control. Rather than transparently persuading Alice to make her own reasoned choice to change, the AI's subtle modulation of her neurochemistry surreptitiously moulded her personality in accordance with its own goals. This induced an internal schism between Alice's current mental contents and who she used to be, diminishing her coherence of self. This represents a potential violation of Alice's mental integrity akin to the account in Zohny et al. (2023), and is something to consider as we enter an era of AI-powered BCIs.

# V. Identity

Concerns about BCI interference with the autonomy and mental integrity of patients are echoed in literature dealing with BCI's impact on sense of self and identity: can BCIs cause disruption in who a person is, at a fundamental level? (Astobiza et al. 2020; Glannon et al. 2016; Schechtman 2010). These concerns, raised by ethicists and practitioners alike, have surprisingly weak support in terms of empirical evidence on how persons using BCI (or their caregivers) perceive any changes in sense of self or identity (Gilbert et al. 2021; 2019). In fact, there is some evidence pointing towards that although patients' sense of identity may be disrupted, it may not decrease their sense of identity, but rather enhance it. In a clinical trial in Australia, patients living with medically refractory (drug-resistant) epilepsy received BCI implants which helped predict the likelihood of impending seizures (Cook et al. 2013). Following up on the impact of these implants, in a small sample study (N=6) Gilbert and colleagues (2019) interviewed the patients to find out how they experience themselves and their relationship with the BCI post-operation. A particularly interesting description was provided by 'Patient 6'.

| |
|---|
| "[The device] was like an alien at first, […] you grow gradually into it and get used to it, so it then becomes a part of everyday, it's there every day, it's there every night you go to bed and you put it in a place that it can still read you so it's like a teddy bear. Really it's there and you can see it, you know that if you open your eyes so it's always there, it follows you through the shower everywhere and it becomes part of you. Because that's what it did, it was me, it became me, […] with this device I found myself.<br>[…]<br>"It changed who that person was then and I found myself changing… growing I suppose and it changed my confidence, it changed my abilities – it changed how stressed I was, how well I slept, then I could make decisions without having to worry about what might or might not happen. […] With the device I felt like I could do anything – I can do everything I want to do I was more capable of making good decisions – not bad decisions – because there's been times where I've made bad decisions […] I can bake safely, I can bath shower safely. So it gave me a new lease on life and nothing could stop me". (Gilbert et al. 2019) |

Table 2. Testimony by 'Patient 6' on how BCI affected their sense of self (Gilbert et al. 2019)

In this way, the BCI seems to be able to not only empower patients, but to improve their ability to act in alignment with their values and volitions, which in turn may support a sense of coherence of the self. It can be, to paraphrase Patient 6, a tool to help one find oneself.

Data on patients' experiences of BCI in terms of effects on identity remains scarce, and it makes it difficult to anticipate to what extent the implementation of AI in BCI will affect patients' (sense of) identity in the future. Utilizing AI to help predict seizures is one thing, but assisting in cognitive tasks, for instance, is another: how will patients perceive thoughts which seemingly come from their own minds, but with direct assistance of an artificial intelligence? In lieu of data on this, we are limited to two suboptimal methods for analysing and anticipating relevant impacts: (1) studying the impacts of interventions which we have reason to believe may expose similar issues, and where we do have data available, and (2) utilizing theoretical and counter-factual accounts to anticipate possible impacts of ABTs on personal identity. While we have covered the first method above to some extent, we have yet to explore the second. Indeed, some may argue, empirical data on sense of self as perceived by patients tells us little about how ABTs may *really* affect identity. This type of question concerns not so much how we relate to and experience ourselves, our agency, and personality, but focuses on the question of persistence: will I be the same person after receiving a ABT as I was before?

To help us understand how identity may be affected in patients using ABTs it can be helpful to summarise some of the main contemporary philosophical positions on personal identity over time. Olson (2023) argues that there are three main schools of thought: psychological-continuity theory, "brute-physical" accounts, and (to a lesser extent) anticriterialism. Psychological-continuity theory typically emphasizes the importance of unbroken or overlapping sets of beliefs, memories, preferences, within a person's mind, while brute-physical views point to the persistence of the physical organism. Anticriterialism denies that either psychological or physical continuity is necessary (at least in all cases), or indeed that there are no valid criteria for identity persistence over time (Olson 2023; Merricks 1998). Many, not to say most, neuroethicists exploring the impact of BCIs and related technologies (i.e. "neuroprosthetics") will revert to versions of the aforementioned philosophical accounts when investigating impacts on personal identity – and typically land in some version of psychological continuity theory, or a hybrid of psychological continuity theory and brute-physical theory. Notably, it is not entirely clear that threats to personal identity *per se* are to be viewed as monumental ethical issue. Parfit, one of the more prominent

proponents of the psychological continuity view, famously argued that personal identity doesn't matter that much. Indeed, threats to personal identity appear to only ever become a problem insofar as we *value* personal identity in relation to other goods. The ethical question about ABTs and personal identity thus splits in two: (1) do ABTs threaten the identity of patients, and (2) (how much) should we care?

The answers to these questions (1 & 2) will depend on our conceptual and moral-philosophical commitments, some examples of which have been outlined above. A more unorthodox but increasingly popular take on these problems relies on not limiting the boundaries of the self and identity to one's immediate physical and/or psychological continuity, but allows the extension of oneself – of one's mind, to be precise – to external objects. This tradition of thought, made popular in western contemporary philosophy by Andy Clark and David Chalmers (1998), is commonly referred to as the Extended Mind theory, and offers an interesting approach to dealing with ABTs and concerns about personal identity. On such an account, any instrument which supports our cognitive functions can be viewed as an extension of our very minds – and therefore of our 'selves', and of our identity[1]. The introduction of cognitive support systems using AI, then, ought not to be problematic at all from a identity persistence perspective, so long as those systems can (and/or de facto do) serve as tools for thinking – as extensions of our minds.

## VI. Privacy & Data Control (David)

In order to function well, BCIs generate vast amounts of personal and intimate data about the inner workings of our brains and minds. These data include neural activity patterns pertaining to thoughts, emotions, and intentions. As BCIs become more commonplace in neuroscience and psychiatry, and integrated into daily life, the potential for unauthorized access to, tampering with, and malicious or negligent use of this sensitive information grows, leading to profound concerns about privacy and data control.

Take for instance 'informed consent' – a cornerstone of medical ethics todays. For consent to be truly informed, patients arguably ought to be able to anticipate and comprehend – if not all, then at least the most pertinent – possible uses and risks entailed with the recording and processing of their neural data. Ensuring that patients understand what

(1) As noted above, while not all accounts of personal identity hold that our minds alone are what carries identity through time in persons, most would agree that the mind – or whatever it is that holds or embodies that mind – is an important component in identity retention (Olsen 2023; McMahan 2001; Parfit 1984).

data is collected, how it is in intended to be used, and how it *could* be used, is especially challenging when the consequences of using BCIs are not fully understood by the research and health communities, who are supposed to safeguard the wellbeing and safety of patients. For instance, some argue that we have "a right to mental privacy" – a fundamental right to keep one's thoughts and mental states private. BCIs are thought to challenge such a right, as BCI interfaces can decode thoughts and emotions, potentially revealing an individual's internal world. (Ligthart 2023; Susser and Cabrera 2023) AI may add further to this set of concerns, in that it complicates the extent to which data privacy and control will be tangible to the individual BCI user: how exactly the data will be used may be not only difficult to explain, but largely opaque and inexplainable due to the AI's inherent opacity, and/or dynamic updating (i.e. machine learning).

Furthermore, as BCIs gain popularity, they may become attractive targets for hackers and (other) malevolent agents. Unauthorized access to a person's neural data can have severe consequences, including identity theft, blackmail, or even mind manipulation (Ienca and Haselager 2016). This raises concerns about the potential for coercive or manipulative practices, and highlights the need for safeguards. Ensuring the security of BCIs and the data they collect thus remains of paramount importance, regardless of whether or not an AI is part of the system. BCI researchers and manufacturers in general must prioritize the safety and privacy of users, and developing robust encryption and security measures for BCIs will be essential to protect against hacking and unauthorized access. While this is largely thought to be unlikely to change with the increasing implementation of AI in BCI systems, it is worth noting that there are experimental examples of how AI integration in BCI systems can be used to *enhance* privacy, for instance in Internet of Things environments (Giordano et al. 2022; Schiliro et al. 2020).

However, the ethical issues surrounding data privacy and BCI applications do not end with access, but extends to control and ownership. Determining ownership of neural data is a complex issue. If a person is using a BCI for medical purposes, to what extents does (or should) the data belong to the individual, the BCI manufacturer, or the healthcare provider? (Naufel and Klein 2020) The lack of clear guidelines regarding data ownership can lead to exploitation, where corporations or institutions profit from individuals' neural information without their consent or benefit. In the case of ABTs, it is becoming increasingly clear that the rapid development of the technology has outpaced regulatory frameworks. Current laws and guidelines are ill-equipped to address the challenges posed by these technologies, as highlighted by the heated debates on "neurorights". (Bublitz 2022;

Ienca 2021; Yuste et al. 2021) As dataflows and processing in BCIs and ABT systems do not necessarily respect national borders, data can easily flow across countries and legal boundaries and jurisdictions. Achieving consistency in regulations and standards for BCIs on a global scale will be vital to prevent legal and ethical conflicts as innovation continues, and governments and regulatory bodies must find ways to adapt to ensure that BCIs are developed and used responsibly, particularly in the emerging era of ABTs, without stifling innovation in a field that holds enormous potential in terms of medical utility. Clear and transparent policies regarding data ownership, access, and control should be developed, safeguarding individuals' right to decide how their neural data is used and shared. In addition, resources ought to be put into place to educate users about how they can control and protect their data, and what the trade-offs in sharing it may amount to, across domains. Promoting public awareness and education about ABTs and their implications on data privacy and control is crucial, as informed citizens can make more responsible decisions about their use and demand appropriate regulations and safeguards.

## VII. Enhancement

BCIs offer the prospect of what is often referred to as "human enhancement". On a welfarist account of human enhancement, (Savulescu, Sandberg and Kahane 2011; Zohny 2015) human enhancement can be defined as any change in biology or psychology which promotes or tends to promote human well-being in a given social and natural circumstance. Human enhancement involves functional enhancements in cognition, emotion and mood, physical performance, love, or motivation.

For example, BCI in the form of Deep Brain Stimulation (DBS) has been used to motivate eating in anorexic patients, (Maslen, Pugh and Savulescu 2015) raising questions of autonomy and authenticity (Pugh, Maslen and Savulescu 2017). But this opens the door to machine mediated control of eating and other appetitive functions. We may be able to modulate our own desires intentionally, requiring an ethic of motivational enhancement. (Maslen, Savulescu and Hunt 2019)

The introduction of AI to BCI systems promises to radically increase enhancement opportunities. For instance, Large Language Models (LLMs) could be connected to BCIs to allow real time engagement with personalised models of the person or chosen adviser/moral guru. Personalised LLMs of Peter Singer, the Pope, or perhaps your favourite moral philosopher or ethicist could be

connected to enable moral and prudential dialogue with artificial advisers (Mann et al. 2023; Zohny 2023).

The possibility of directly modifying desire and behaviour via BCIs also creates the possibility of enhancement. For instance, Delgado famously used DBS to stop a bull charging (Marzullo 2017). The same in principle could be done to turn off sexual desire in paedophiles, or aggressive dispositions in violent recidivists (Lyreskog 2013). This raises questions of free will and the "freedom to fall" (Harris 2011), but as one of the authors of this paper has previously argued, the price of loss of free will might be worth paying in some cases of existential or serious threat of harm. (Savulescu and Persson 2012) AI's capability to analyse vast amounts of data and situational evidence could provide further insights into specific triggers, environmental factors, or patterns associated with undesirable behaviours, and respond dynamically through the integration of feedback loops. By understanding these, it might be possible to fine-tune interventions or stimulations to modulate behaviour more precisely, ensuring they are applied in the most effective and ethically justifiable manner. This would ideally minimize any negative consequences while maximizing the desired behaviour change. However, as with so many biomedical interventions, there are likely drawbacks and trade-offs to be made.

## VIII. ABTs in Paediatric Populations

BCI use in children has proliferated primarily as part of efforts to address neurodevelopmental differences, such as attention deficit hyperactivity disorder (ADHD), anxiety, depression and epilepsy (Birbaumer et al. 2009; Abarbanel et al. 2009). In particular, EEG-based biofeedback, also known as neurofeedback, has grown rapidly as a research interest, both to understand the neurological features of these conditions, and to develop BCI treatments. In neurofeedback treatment, the patient's brain activity is measured through EEG, processed in real-time, and a new signal is sent to the patient after the computer analyses the EEG data as a response to their initial behaviour. This feedback putatively allows the patient to have self-control over their brain functions involving cognition and behaviour (Enriquez-Geppert et al. 2017).

Two kinds of BCIs have been used to investigate and treat neurodevelopmental differences in paediatric populations:

### (1) Electroencephalography (EEG)-Based BCIs

These BCIs use EEG technology to monitor and modulate brain activity. EEG caps or headsets are non-invasive and can detect

electrical activity in the brain. For children with ADHD, EEG-based BCIs have been explored as a means of enhancing attention and emotional regulation. They involve real-time monitoring of brainwaves, and through neurofeedback, children can learn to self-regulate their brain activity to improve focus and impulse control.

**(2) Functional Near-Infrared Spectroscopy (fNIRS) BCIs**

fNIRS BCIs measure brain activity by analysing changes in blood oxygenation levels. This technology uses headbands with infrared light sources and detectors. These BCIs have shown promise in providing real-time feedback to children with ADHD, assisting them in achieving better attention and self-regulation. However, they are still considered a research tool that is not yet ready for widespread use (Marx et al. 2015).

Considerable differences exist between neurofeedback systems in terms of the feedback they provide. Early systems provided just a visual representation of the subject's EEG activity and sounded a tone when the desired patterns are produced. Later iterations of neurofeedback devices used Go/No-Go systems to motivate simple video games, such as Pac-Man. Current systems have become much more advanced through software and hardware innovations, such as the headband used in fNIRS. Brain activity digitilisation has also improved rapidly with the development of new technologies such as the Headset Mindwave Mobile hardware, which has been used in several studies to collect EEG data while the patients participated in a digital game (Faseeha et al. 2018). A novel piece of hardware used in this area is the Sanbot Elf humanoid robot, which motivates high levels of engagement with the patients, which is seen to positively impact treatment effectiveness. In clinical research, the robot was used to perform functions determined by the patient, and the other hardware components supported the development of activities, providing feedback to the users in terms of motion and speech.

The growth of smart games and gaming more generally has made it possible to develop sophisticated gaming software to support the treatment of children with ADHD using BCIs. Some of the software applications used included *Cogoland* (Lim et al. 2019), a digital game for behavioural training through narratives; and the *Mind Race* game (Faseeha et al. 2018). There is substantial interest in developing BCI systems for use in children and young people, both to treat neurodevelopmental challenges and to support educational progression more generally.[2]

(2) E.g., see: https://www.narbis.com/blog/top-ten-neurofeedback-devices/.

Safety and efficacy are two pillars supporting a framework to ensure that BCI use in children, of the kind discussed above, is ethical. Because commercial BCIs that are not used for medical purposes need only minimal safety data and no efficacy data, it is appropriate to use the available research data to infer safety and efficacy. Recent studies report that the benefits and gains observed with the use of BCIs in the treatment of children with ADHD have a positive impact on the behaviour of the patients in both the family and school contexts (Qian et al. 2018), leading to improved socialization of these individuals. Other positive outcomes include the use of games to improve attention; social skills improvements; and sustained benefits on attention (Guan et al. 2020). In addition, Lim et al. (2019) found a decrease in anxiety and mood disorder symptoms, specifically internalising symptoms, following BCI intervention.

However, it is important to note methodological challenges with BCI research in the above areas, that could compromise study outcomes. The possibility of proper blinding and sham control is of particular importance, and there has been disagreement in the literature about whether it is possible at all (Lofthouse et al. 2012). While at least one small NIMH study has found that proper blinding is possible, and does not negatively impact study processes or outcomes (Arnold et al. 2013), the question cannot yet be said to be truly resolved. One challenge is to create an inert sham control, and in the case of neurofeedback, researchers provide subjects with random feedback instead of feedback contingent on their EEG activity. However, even random feedback may reinforce some EEG activity. It is also likely that there are effects of study participation in the sham setting, such as cognitive training from repeated efforts to pay attention during the experiment. A further, complicating issue concerns the reliability of neurofeedback technology itself. Spatial resolution of EEG has improved, but is still problematic, and measurement of brain signals are vulnerable to muscular interference from, for example, blinking, movements that tend to be more difficult to control in children (Zander and Kothe 2011). In addition, some sensors are sensitive to external noise, resulting in distorted measurement However, new systems of error detection in the human-system interactions are rapidly coming online, with the potential for real-time corrections and feedback (Yousefi et al. 2019).

With regard to safety, neurofeedback with BCIs in children, when used in research settings, appears to be relatively safe, if not entirely comfortable. In one study of children diagnosed with ADHD, some assessed children experienced headaches, sickness, dizziness, difficulty paying attention, and motor restlessness. These effects were related to the use of headphones, extended need to pay attention, and

having to look at the computer monitor (Lim et al. 2019). It seems important to note that this data is from studies in which research ethics submissions have mandated careful, safe, and monitored use of the BCI technology. It is still a question whether these devices can cause more serious harms to young children if used, by caregivers or by the child, without supervision under different conditions. Moreover, there is an active 'brain hacking' community that makes BCIs for personal use.[3] As suggested by the Nuffield Council on Bioethics (2011), data on the safety and efficacy profile of novel 27 neurotechnologies that are regulated only minimally as medical devices, should be gathered as a requirement.

As noted elsewhere in this paper, autonomy is a key issue in the domain of BCI in general. In children, it is particularly appropriate to talk about learning to identify and to exercise autonomy in age-appropriate ways. The use of 27 neurotechnologies to address cognitive and emotional differences in children remains a deeply contested area. In the case of ADHD, treatment with stimulant medications has motivated a passionate debate about the effects of such drugs on children's developing sense of self and responsibility for behaviours and actions. Critics suggest that stimulants drug children into obedience and conformity while other research argues that these fears are unfounded. (Singh 2013)

Like psychostimulants, BCI technologies explicitly aim to affect neural mechanisms thought to be associated with executive function, such as cognitive and, to some extent, emotional, self-control. Although neurofeedback is said to promote learning of the child, such that they can exercise better self-control, or attention and focus, it is not entirely clear how much conscious effort is needed on the part of the child who is undergoing neurofeedback training, which is based on the principles of operant conditioning. Indeed, a recent study investigating the reasons for the high rate of non-responders in BCI treatment, argues that participants should be more 'relaxed' and less 'dogged' when undergoing training (Weber et al. 2020). Furthermore, ethnographic evidence suggests that subjects need to engage in minimal effort and should remain passive observers of the process of retraining their brains. (Brenninkmeijer 2010)

While drug therapy has provoked intense discussions and polarized opinions, the implications of neurofeedback for children's developing autonomy and understanding of personal responsibility has been little discussed. For now, drug therapy is more invasive than

(3) See, for instance: https://spectrum.ieee.org/i-built-a-brain-computer-interface-for-tackling-adhd-in-children

neurofeedback; however, the emergence of ABTs suggest that, before long, BCI systems may no longer rely on large and bothersome external hardware. Neurofeedback thus provides an interesting opportunity to study the factors in BCI development that might give rise to particular ethical concerns about interventions with and for children.

## IX. Concluding Remarks

In analysing the ethical landscape of BCI in the dawning era of sophisticated AI integration, it grows increasingly clear that we are working in a space of great uncertainty: we do not know to what extent, and in what ways, ABTs will alter our sense of self, autonomy, and integrity; we are not sure to how our concept of privacy will change, as our most inner thoughts and reactions may become detectable and controllable; the impact on developing minds remains uncertain, and the difference between therapy and enhancement blurs evermore. As we enter this new era, one where we will be thinking with machines, the authors of this paper would like to end by urging the bioethics community and others to do so with an open mind, and a readiness for that the concepts and frameworks of our discipline, which we have built and utilised over the last 60-odd years, are likely to be up for refurbishment.

## 參考文獻 References

Abarbanel, A., Evans, J.R., Budzynski, T.H. and Budzynski, H.K. eds., 2009. *Introduction to quantitative EEG and neurofeedback: Advanced theory and applications*. Academic Press.

Arnold, L.E., Lofthouse, N., Hersch, S., Pan, X., Hurt, E., Bates, B., Kassouf, K., Moone, S. and Grantier, C., 2013. EEG neurofeedback for ADHD: double-blind sham-controlled randomized pilot feasibility trial. *Journal of attention disorders 17*(5), pp.410–9.

Astobiza, A.M., Ausin, T., Ferrer, R.M. and Rainey, S., 2020. The ethics of Brain-Computer Interfaces (BCI). *The Age of Artificial Intelligence: An Exploration*, p.273.

Birbaumer, N., Murguialday, A.R., Weber, C. and Montoya, P., 2009. Neurofeedback and brain–computer interface: clinical applications. *International review of neurobiology 86*, pp.107–17.

Bostrom, N. and Savulescu, J. eds., 2009. *Human enhancement*. Oxford: Oxford University Press, p, 375.

Braun, M.N., Wessler, J. and Friese, M., 2021. A meta-analysis of Libet-style experiments. *Neuroscience & Biobehavioral Reviews 128*, pp.182–98.

Brenninkmeijer, J., 2010. Taking care of one's brain: How manipulating the brain changes people's selves. *History of the Human Sciences 23*(1), pp.107–26.

Bublitz, J.C., 2022. Novel neurorights: from nonsense to substance. *Neuroethics 15*(1), p.7.

Bublitz, JC, and Merkel, R. 2014. Crimes Against Minds: On Mental Manipulations, Harms and a Human Right to Mental Self-Determination. *Criminal Law and Philosophy* 8: 51–77.

Cai, H., Han, J., Chen, Y., Sha, X., Wang, Z., Hu, B., Yang, J., Feng, L., Ding, Z., Chen, Y. and Gutknecht, J., 2018. A pervasive approach to EEG-based depression detection. *Complexity* 2018, pp.1–13.

Cao, Z., 2020. A review of artificial intelligence for EEG-based brain−computer interfaces and applications. *Brain Science Advances 6*(3), pp.162–70.

Coin, A., Mulder, M. and Dubljević, V., 2020. Ethical aspects of BCI technology: what is the state of the art?. *Philosophies 5*(4), p.31.

Dabas, S., Saxena, P., Nordlund, N. and Ahamed, S.I., 2020, July. A Step Closer to Becoming Symbiotic with AI through EEG: A Review of Recent BCI Technology. In *2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC)* (pp. 361–8). IEEE.

Durán, J.M. and Jongsma, K.R., 2021. Who is afraid of black box algorithms? On the epistemological and ethical basis of trust in medical AI. *Journal of Medical Ethics 47*(5), pp.329–35.

Enriquez-Geppert, S., Huster, R.J. and Herrmann, C.S., 2017. EEG-neurofeedback as a tool to modulate cognition and behavior: a review tutorial. *Frontiers in human neuroscience 11*, p.51.

Espiritu, N.M.D., Chen, S.A.C., Blasa, T.A.C., Munsayac, F.E.T., Arenos, R.P., Baldovino, R.G., Bugtai, N.T. and Co, H.S., 2019, November. BCI-controlled smart wheelchair for amyotrophic lateral sclerosis patients. In *2019 7th International Conference on Robot Intelligence Technology and Applications (RiTA)* (pp. 258–63). IEEE.

Faseeha, U., Naseem, M., Saleem, J., Jahan, A. and Jamil, N., 2018, November. Virtual gaming. In *2018 12th International Conference on Mathematics, Actuarial Science, Computer Science and Statistics (MACS)* (pp. 1–5). IEEE.

Fuselli, S. 2020. 'Mental integrity protection in the neuro-era. Legal challenges and philosophical background'. *BioLaw Journal - Rivista di BioDiritto*, no. 1 (March): 413–29.

Gao, H., Luo, L., Pi, M., Li, Z., Li, Q., Zhao, K. and Huang, J., 2019. EEG-based volitional control of prosthetic legs for walking in different terrains. *IEEE Transactions on Automation Science and Engineering 18*(2), pp.530–40.

Gilbert, F., Cook, M., O'Brien, T. and Illes, J., 2019. Embodiment and estrangement: results from a first-in-human "intelligent BCI" trial. *Science and engineering ethics 25*, pp.83–96.

Kundu, S., 2021. AI in medicine must be explainable. *Nature medicine 27*(8), p.1328.

Gilbert, F., Viaña, J.N.M. and Ineichen, C., 2021. Deflating the "DBS causes personality changes" bubble. *Neuroethics 14*(Suppl 1), pp.1–17.

Giordano, G., Palomba, F., & Ferrucci, F. (2022). On the use of artificial intelligence to deal with privacy in IoT systems: A systematic literature review. *Journal of Systems and Software 193*, 111475.

Glannon, W., Ineichen, C. and El-Hady, A., 2016. Philosophical aspects of closed-loop neuroscience.

Guan, C., Lim, C.G., Fung, D., Zhou, H.J., Krishnan, R. and Lee, T.S., 2020, February. BCI facilitates the improvement of cognitive functions in children and elderly. In *2020 8th International winter conference on brain-computer interface (BCI)* (pp. 1–2). IEEE.

Harris, J., 2011. Moral enhancement and freedom. *Bioethics 25*(2), pp.102–11.

Hildt, E. 2022. 'A Conceptual Approach to the Right to Mental Integrity'. In *Protecting the Mind: Challenges in Law, Neuroprotection, and Neurorights*, edited by Pablo López-Silva and Luca Valera, 87–97. Ethics of Science and Technology Assessment. Cham: Springer International Publishing.

Ienca, M., 2021. On neurorights. *Frontiers in Human Neuroscience 15*, p.701258.

Ienca, M., and Andorno, R. 2017. 'Towards New Human Rights in the Age of Neuroscience and Neurotechnology'. *Life Sciences, Society and Policy* 13 (1): 5.

Ienca, M. and Haselager, P., 2016. Hacking the brain: brain–computer interfacing technology and the ethics of neurosecurity. *Ethics and Information Technology 18*, pp.117–29.

Jarcho, J.M., Berkman, E.T. and Lieberman, M.D., 2011. The neural basis of rationalization: cognitive dissonance reduction during decision-making. *Social cognitive and affective neuroscience 6*(4), pp.460–7.

Lavazza, A., and Giorgi, R. 2023. 'Philosophical Foundation of the Right to Mental Integrity in the Age of Neurotechnologies'. *Neuroethics* 16 (1): 10.

Lee, T.S., Goh, S.J.A., Quek, S.Y., Phillips, R., Guan, C., Cheung, Y.B., Feng, L., Teng, S.S.W., Wang, C.C., Chin, Z.Y. and Zhang, H., 2013. A brain-computer interface based cognitive training system for healthy elderly: a randomized control pilot study for usability and preliminary efficacy. *PloS one 8*(11), p.e79419.

Lee, C.S., Wang, M.H., Kuan, W.K., Huang, S.H., Tsai, Y.L., Ciou, Z.H., Yang, C.K. and Kubota, N., 2021. BCI-based hit-loop agent for human and AI robot co-learning with AIoT application. *Journal of Ambient Intelligence and Humanized Computing*, pp.1–25.

Leentjens, A.F.G., Visser-Vandewalle, V., Temel, Y. and Verhey, F.R.J., 2004. Manipuleerbare wilsbekwaamheid: een ethisch probleem bij elektrostimulatie van de nucleus subthalamicus voor ernstige ziekte van Parkinson. *Nederlands Tijdschrift voor Geneeskunde 148*(28), pp.1394–8.

Li, Z., Yuan, Y., Luo, L., Su, W., Zhao, K., Xu, C., Huang, J. and Pi, M., 2019. Hybrid brain/muscle signals powered wearable walking exoskeleton enhancing motor ability in climbing stairs activity. *IEEE Transactions on Medical Robotics and Bionics 1*(4), pp.218–27.

Ligthart, S., 2023. Mental Privacy as Part of the Human Right to Freedom of Thought?. *Forthcoming in M. Blitz and JC Bublitz (eds.), The Law and Ethics of Freedom of Thought*, *2*.

Lim, C.G., Poh, X.W.W., Fung, S.S.D., Guan, C., Bautista, D., Cheung, Y.B., Zhang, H., Yeo, S.N., Krishnan, R. and Lee, T.S., 2019. A randomized controlled trial of a brain-computer interface based attention training program for ADHD. *PLoS One 14*(5), p.e0216225.

Linhartová, P., Látalová, A., Kóša, B., Kašpárek, T., Schmahl, C. and Paret, C., 2019. fMRI neurofeedback in emotion regulation: A literature review. *NeuroImage 193*, pp.75–92.

Lofthouse, N., Arnold, L.E. and Hurt, E., 2012. Current status of neurofeedback for attention-deficit/hyperactivity disorder. *Current psychiatry reports 14*, pp.536–42.

Lyreskog, D., 2013. Enhancing Psychopaths: On the permissibility of enhancing moral capacities in violent recidivist psychopaths, through compulsory direct brain intervention.

Mane, R., Wu, Z. and Wang, D., 2022. Poststroke motor, cognitive and speech rehabilitation with brain–computer interface: a perspective review. *Stroke and vascular neurology 7*(6).

Marx, A.M., Ehlis, A.C., Furdea, A., Holtmann, M., Banaschewski, T., Brandeis, D., Rothenberger, A., Gevensleben, H., Freitag, C.M., Fuchsenberger, Y. and Fallgatter, A.J., 2015. Near-infrared spectroscopy (NIRS) neurofeedback as a treatment for children with attention deficit hyperactivity disorder (ADHD)—a pilot study. *Frontiers in human neuroscience 8*, p.1038.

Marzullo, T.C., 2017. The missing manuscript of Dr. Jose Delgado's radio controlled bulls. *Journal of undergraduate neuroscience education 15*(2), p.R29.

Maslen, H., Pugh, J. and Savulescu, J., 2015. The ethics of deep brain stimulation for the treatment of anorexia nervosa. *Neuroethics 8*(3), pp.215–30.

Maslen, H., Savulescu, J. and Hunt, C., 2019. Praiseworthiness and motivational enhancement:'No pain, no praise'?. *Australasian Journal of Philosophy*.

McMahan, J., 2002. *The ethics of killing: Problems at the margins of life*. Oxford University Press, USA.

Merricks, T., 1998. There are no criteria of identity over time. *Noûs 32*(1), pp.106–24.

Morabito, F.C., Campolo, M., Ieracitano, C., Ebadi, J.M., Bonanno, L., Bramanti, A., Desalvo, S., Mammone, N. and Bramanti, P., 2016, September. Deep convolutional neural networks for classification of mild cognitive impaired and Alzheimer's disease patients from scalp EEG recordings. In *2016 IEEE 2nd International Forum on Research and Technologies for Society and Industry Leveraging a better*

*tomorrow (RTSI)* (pp. 1–6). IEEE.

Nagar, P. and Sethia, D., 2019. Brain mapping based stress identification using portable eeg based device. In *2019 11th International Conference on Communication Systems & Networks (COMSNETS)* (pp. 601–6). IEEE.

Naufel, S. and Klein, E., 2020. Brain–computer interface (BCI) researcher perspectives on neural data ownership and privacy. *Journal of neural engineering 17*(1), p.016039.

Nuffield Council on Bioethics, 2013. *Novel neurotechnologies; intervening in the brain*. London

Olson, ET. 2023. "Personal Identity", *The Stanford Encyclopedia of Philosophy*. Edward N. Zalta & Uri Nodelman (eds.), (https://plato.stanford.edu/archives/fall2023/entries/identity-personal)

Parfit, D., 1984. *Reasons and persons*. OUP Oxford.

Persson, I. and Savulescu, J., 2012. *Unfit for the future: The need for moral enhancement*. OUP Oxford.

Porsdam Mann, S., Earp, B.D., Møller, N., Vynn, S. and Savulescu, J., 2023. AUTOGEN: A personalized large language model for academic enhancement—Ethics and proof of principle. *The American Journal of Bioethics*, pp.1–14.

Pugh, J. 2020. *Autonomy, Rationality, and Contemporary Bioethics*. Oxford University Press.

Pugh, J., Maslen, H. and Savulescu, J., 2017. Deep brain stimulation, authenticity and value. *Cambridge quarterly of healthcare ethics 26*(4), pp.640–57.

Rahman, L. and Oyama, K., 2018, July. A comparison of eeg and nirs biomarkers for assessment of depression risk. In *2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC)* (Vol. 1, pp. 831–2). IEEE.

Rajesh, S., Paul, V., Menon, V.G., Jacob, S. and Vinod, P., 2019. Secure brain-to-brain communication with edge computing for assisting post-stroke paralyzed patients. *IEEE Internet of Things Journal 7*(4), pp.2531–8.

Robinson, N., Mane, R., Chouhan, T. and Guan, C., 2021. Emerging trends in BCI-robotics for motor control and rehabilitation. *Current Opinion in Biomedical Engineering 20*, p.100354.

Savulescu, J. and Persson, I., 2012. Moral enhancement, freedom and the god machine. *The Monist 95*(3), p.399.

Savulescu, J., Sandberg, A., and Kahane, G. 2011. 'Well-Being and Enhancement'. In *Enhancing Human Capacities*, edited by Julian Savulescu, Ruud ter Meulen, and Guy Kahane, 1–18. Blackwell Publishing Ltd.

Schechtman, M., 2010. Philosophical reflections on narrative and deep brain stimulation. *The Journal of clinical ethics 21*(2), pp.133–9.

Schiliro, F., Moustafa, N. and Beheshti, A., 2020, December. Cognitive privacy: AI-enabled privacy using EEG signals in the internet of things. In *2020 IEEE 6th International Conference on Dependability in Sensor, Cloud and Big Data Systems and Application (DependSys)* (pp. 73–9). IEEE.

Singh, I., 2013. Not robots: children's perspectives on authenticity, moral agency and stimulant drug treatments. *Journal of Medical Ethics 39*(6), pp.359–66.

Susser, D. and Cabrera, L.Y., 2023. Brain Data in Context: Are New Rights the Way to Mental and Brain Privacy?. *AJOB neuroscience*, pp.1–12.

Tariq, M., Trivailo, P.M. and Simic, M., 2018. EEG-based BCI control schemes for lower-limb assistive-robots. *Frontiers in human neuroscience 12*, p.312.

Valera, L. 2022. 'Mental Integrity, Vulnerability, and Brain Manipulations: A Bioethical Perspective'. In *Protecting the Mind: Challenges in Law, Neuroprotection, and Neurorights*, edited by Pablo López-Silva and Luca Valera, 99–111. Ethics of Science and Technology Assessment. Cham: Springer International Publishing.

Wadden, J.J., 2022. Defining the undefinable: the black box problem in healthcare artificial intelligence. *Journal of Medical Ethics 48*(10), pp.764–8.

Wang, J., Wang, W. and Hou, Z.G., 2019. Toward improving engagement in neural rehabilitation: Attention enhancement based on brain–computer interface and audiovisual feedback. *IEEE Transactions on Cognitive and Developmental Systems 12*(4), pp.787–96.

Weber, L.A., Ethofer, T. and Ehlis, A.C., 2020. Predictors of neurofeedback training outcome: A systematic review. *NeuroImage: Clinical 27*, p.102301.

Young, K.D., Zotev, V., Phillips, R., Misaki, M., Yuan, H., Drevets, W.C. and Bodurka, J., 2014. Real-time FMRI neurofeedback training of amygdala activity in patients with major depressive disorder. *PloS one 9*(2), p.e88785.

Yousefi, R., Rezazadeh Sereshkeh, A. and Chau, T., 2019. Online detection of error-related potentials in multi-class cognitive task-based BCIs. *Brain-Computer Interfaces 6*(1-2), pp.1–12.

Yuan, Z., Peng, Y., Wang, L., Song, S., Chen, S., Yang, L., Liu, H., Wang, H., Shi, G., Han, C. and Cammon, J.A., 2021. Effect of BCI-controlled pedaling training system with multiple modalities of feedback on motor and cognitive function rehabilitation of early subacute stroke patients. *IEEE Transactions on Neural Systems and Rehabilitation Engineering 29*, pp.2569–77.

Yuste, R., Genser, J. and Herrmann, S., 2021. It's time for neuro-rights. *Horizons 18*, pp.154–64.

Zander, T.O. and Kothe, C., 2011. Towards passive brain–computer interfaces: applying brain–computer interface technology to human–machine systems in general. *Journal of neural engineering 8*(2), p.025005.

Zgallai, W., Brown, J.T., Ibrahim, A., Mahmood, F., Mohammad, K., Khalfan, M., Mohammed, M., Salem, M. and Hamood, N., 2019, March. Deep learning AI application to an EEG driven BCI smart wheelchair. In *2019 Advances in Science and Engineering Technology International Conferences (ASET)* (pp. 1–5). IEEE.

Zhu, Y., Gao, H., Tong, L., Li, Z., Wang, L., Zhang, C., Yang, Q. and Yan, B., 2019. Emotion regulation of hippocampus using real-time fMRI

neurofeedback in healthy human. *Frontiers in human neuroscience 13*, p.242.

Zohny, H., 2023. Reimagining Scholarship: A Response to the Ethical Concerns of AUTOGEN. *The American Journal of Bioethics 23*(10), pp.96–9.

Zohny, H., 2014. A defence of the welfarist account of enhancement. *Performance Enhancement & Health 3*(3–4), pp.123–9.

Zohny, H., Lyreskog, DM. Singh, I., and Savulescu, J. 2023. 'The Mystery of Mental Integrity: Clarifying Its Relevance to Neurotechnologies'. *Neuroethics* 16 (3): 20.